

R E P O R T R E S U M E S

ED 013 574

FL 000 379

MLA FOREIGN LANGUAGE PROFICIENCY TESTS FOR TEACHERS AND
ADVANCED STUDENTS.

BY- STARR, WILMARTH H.

MODERN LANGUAGE ASSN. OF AMERICA, NEW YORK, N.Y.

EDRS PRICE MF-\$0.25 HC-\$0.60 15P.

PUB DATE SEP 62

DESCRIPTORS- *ADVANCED STUDENTS, *LANGUAGE PROFICIENCY,
*LANGUAGE TESTS, *LANGUAGE TEACHERS, *NATIONAL COMPETENCY
TESTS, FRENCH, GERMAN, ITALIAN, LANGUAGE SKILLS, RUSSIAN,
SPANISH, STANDARDIZED TESTS, TEST RELIABILITY, TEST VALIDITY,
TESTING PROGRAMS, TEACHER QUALIFICATIONS, MLA PROFICIENCY
TESTS, EDUCATIONAL TESTING SERVICE,

THE DEVELOPMENT AND EVALUATION OF THE MODERN LANGUAGE
ASSOCIATION (MLA) FOREIGN LANGUAGE PROFICIENCY TESTS FOR
TEACHERS AND ADVANCED STUDENTS ARE THE SUBJECTS OF THIS FINAL
PROJECT REPORT. FOLLOWING AN ACCOUNT OF THE EVENTS THAT LED
TO THE AWARDED OF A GOVERNMENT CONTRACT TO MLA TO DEVELOP
NATIONALLY STANDARDIZED QUALIFICATION TESTS AND A DESCRIPTION
OF THE LOGISTIC AND EVALUATION PROBLEMS INVOLVED IN
DEVELOPING THESE TWO 31-TEST BATTERIES IN FRENCH, GERMAN,
ITALIAN, RUSSIAN, AND SPANISH, THE REPORT APPRAISES, WITH
SUPPORTING STATISTICAL DATA, THE HIGH LEVEL OF RELIABILITY
AND STATISTICAL VALIDITY OF THESE TESTS IN THE COMPETENCY
AREAS OF LISTENING COMPREHENSION, SPEAKING, READING, WRITING,
APPLIED LINGUISTICS, CULTURE AND CIVILIZATION, AND
PROFESSIONAL PREPARATION. A COMPARISON OF PRE-TEST AND
POST-TEST NORMS FURNISHES THE GOVERNMENT AND THE PROFESSION
WITH OBJECTIVE DATA FOR FUTURE TEST REVISIONS AND FOR SUPPORT
AND ANALYSIS OF INSTITUTE PROGRAMS. FOUR APPENDIXES INCLUDE
INFORMATION ABOUT MODERN FOREIGN LANGUAGE TEACHER
QUALIFICATIONS, A DIRECTORY OF THE MEMBERS OF THE TEST
CONSTRUCTION COMMITTEES, AND DATA JUSTIFYING TEST RELIABILITY
AND THE INTERCORRELATIONS BETWEEN SKILLS. THIS ARTICLE IS A
REPRINT FROM "PMLA," VOLUME 77, NUMBER 4, PART 2, SEPTEMBER
1962. (AB)

ED013574

PMLA

PUBLICATIONS OF THE MODERN LANGUAGE ASSOCIATION OF AMERICA

Issued Five Times a Year

VOLUME LXXVII

SEPTEMBER 1962

NUMBER 4, PART 2

MLA FOREIGN LANGUAGE PROFICIENCY TESTS FOR TEACHERS AND ADVANCED STUDENTS

WILMARTH H. STARR

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

Published by the MODERN LANGUAGE ASSOCIATION OF AMERICA

Materials Center
4 Washington Place
New York, N.Y. 10003

"PERMISSION TO REPRODUCE THIS
COPYRIGHTED MATERIAL HAS BEEN GRANTED
BY PMLA

TO ERIC AND ORGANIZATIONS OPERATING
UNDER AGREEMENTS WITH THE U.S. OFFICE OF
EDUCATION. FURTHER REPRODUCTION OUTSIDE
THE ERIC SYSTEM REQUIRES PERMISSION OF
THE COPYRIGHT OWNER."

FL 000 379

MLA FOREIGN LANGUAGE PROFICIENCY TESTS FOR TEACHERS AND ADVANCED STUDENTS*

BY WILMARTH H. STARR

Director, Modern Language Association Testing Project

I. Brief History of the Project: Since 1952, the Foreign Language Program of the Modern Language Association of America, responding to the national urgency with regard to foreign languages, has been engaged in a vigorous campaign aimed in large part at improving foreign-language teaching in our country.

In 1955, as one of its activities, the Steering Committee of the Foreign Language Program formulated the "Qualifications for Secondary School Teachers of Modern Foreign Languages," a statement which was subsequently endorsed for publication by the MLA Executive Council, by the Modern Language Committee of the Secondary Education Board, by the Committee on the Language Program of the American Council of Learned Societies, and by the executive boards or councils of the following national and regional organizations: National Federation of Modern Language Teachers Associations, American Association of Teachers of French, American Association of Teachers of German, American Association of Teachers of Italian, American Association of Teachers of Spanish and Portuguese, American Association of Teachers of Slavic and East European Languages, Central States Modern Language Teachers Association, Middle States Association of Modern Language Teachers, New England Modern Language Association, Northeast Conference on the Teaching of Foreign Languages, Northwest Conference on Foreign Language Teaching, Philological Association of the Pacific Coast, Rocky Mountain Modern Language Association, South Atlantic Modern Language Association, and South-Central Modern Language Association.

The statement established three general levels of proficiency (Minimal, Good, and Superior) for seven areas of language teaching competencies: 1) aural understanding, 2) speaking, 3) reading, 4) writing, 5) language analysis, 6) culture, 7) professional preparation. In subsequent conferences involving national leaders representing the field of educational administration and national leaders in the foreign language profession the need for the development of standardized proficiency tests as an aspect of teacher preparation and certification was discussed and endorsed. It was obvious to many that the state-

ment of desiderata, no matter how strongly representative of a consensus of the profession, would not be as effective as the situation demanded until nationally standardized tests could be developed that would implement the descriptions of competencies.

In the spring of 1959 the means to develop nationally standardized qualification tests for teachers of foreign languages were implemented under a contract between the U. S. Office of Education (NDEA Title VI) and the Modern Language Association.

Professor Wilmarth H. Starr, then Head of the Department of Foreign Languages and Classics at the University of Maine and currently Head of the All-University Department of Romance and Slavic Languages and Literatures at New York University, was named Project Director. During the summer and early fall of 1959 the plan of organization for the project was developed, a preliminary study of typical existing foreign language tests was carried out, and the first meeting of the Area Committees to refine test objectives and to develop test specifications was held.

It is important to note that the Qualifications Statement referred to above, as modified for use by the Committees (see Appendix A), became the guide for the range and spectrum of the test batteries and has remained so throughout the period of test development. The test batteries thus serve the interests of the profession as originally defined by the profession.

In this respect, it was a basic assumption of the project proposal that the tests should essentially be developed by the people who use them. The tests are then a product of the foreign-language teaching profession as illustrated by the nationally recognized names of the teachers and institutions which have been involved in the project since its inception (see Appendix B). In support of this professional attitude, which from the beginning called upon the best efforts of a nationally representative cross section of foreign-language scholars and teachers within the Mod-

* The final report on Contract No. SAE 8349 submitted to the Acting Director of the Division of College and University Assistance, Bureau of Educational Assistance Programs, United States Office of Education, on 15 June 1962.

ern Language Association at university, college, and secondary-school levels, the MLA has been reinforced in its efforts, also from the beginning, by the constant and close collaboration of Educational Testing Service, whose representatives in each language and area and whose testing experts have added their resources to those of the professional language teacher.

Evidence of the objectivity with which we have approached our task is the way in which the name of the test batteries has been changed during the course of the test development. From the original concept of Qualifications Tests for Secondary School Teachers of Foreign Languages, there has been a series of modifications to Qualifications Tests for Teachers of Modern Foreign Languages to the present title, Foreign Language Proficiency Tests for Teachers and Advanced Students. In the first place it became apparent that the word "qualifications" suggested impingement upon the rights of States and individual institutions to determine their own qualification criteria, since it seemed to imply a tacit assumption that the MLA or the government was undertaking to "impose" qualifications standards. The proper view implicit in the new title is that the test batteries are sensitive instruments to measure *proficiency* as related to national norms, but the presumption of "qualifications" is clearly left to the agencies established for such purposes.

In the second place, it has become equally clear that important uses can be made of the tests in terms of proficiency measurement for advanced students in teacher-training programs, in M.A. and Ph.D. programs, as guides for placement and diagnostic purposes or as indicators of achievement.

Needless to say, the tests themselves have also undergone an elaborate process of evolution. The 62 preliminary tests developed in more than 75 meetings of test construction committees in

1959-60 were considerably longer than necessary for even the high reliability required of them. This was because we did not wish to pre-judge sensitive problems in connection with speaking and writing and because we wished to experiment with a maximum of item types from which we could then select the best on the basis of data analysis. In the summer and fall of 1960 over 30,000 individual preliminary tests were administered and over 26,000 were scored. As a result we were provided with an enormous amount of data for revision purposes and at the same time were able to give significant information about the test population in 37 NDEA Summer Institutes, 5 Academic-Year Institutes, and selected control groups in the Carnegie Inter-University programs and the Middlebury Italian Summer School.

Apart from the considerable logistics problem involved in printing, coding, and shipping more than three tons of test materials to various centers of administration in this country and abroad, major problems were associated with the fact that scoring teams had to be trained not only to use the electronic equipment effectively, but to control the sensitive evaluation processes established for this pioneering venture. It must be pointed out that there was no previous experience to draw from which was predictably applicable to a five-language, oral-production operation on the scale of our project. Similar problems were experienced in connection with the scoring of that part of the writing tests not subject to machine scoring.

In connection with the scoring of tests, which required human scorers, it is worth noting, as a testimonial of professional loyalty, that significant numbers of trained scorers from the 1960 and 1961 scoring sessions will return as the nucleus for the 1962 scoring teams.

In the fall and winter of 1960-61, the data compiled by Educational Testing Service was

Test	Actual Testing Time in minutes Preliminary Forms	Actual Testing Time in minutes Revised Forms	Saving in minutes
Listening Comprehension	40	20	20
Speaking	40	15	25
Reading	70	40	30
Writing	90	45	45
Applied Linguistics	40	40	—
Culture-Civilization	60	30	30
Professional Preparation	70	45	25
	—	—	—
	410=	235=	175=
	6 hrs. 50 min.	3 hrs. 55 min.	2 hrs. 55 min.

explained to and analyzed by the committee chairmen in a series of seven meetings preliminary to final form revisions. As a result, it was possible to develop final batteries of predictably high reliability and discriminatory power with substantially sharpened, refined, and shortened tests. The final forms were ready for administration in the summer of 1961 to the 75 Summer and Academic-Year NDEA Institutes, an additional control group studying in Mexico (University of Arizona) and two groups going to Russia (Indiana University and the University of Michigan).

The preceding table illustrates actual testing time saved by the final forms over the preliminary forms. The length of the revised forms was determined by the number of items necessary to insure figures for reliability and discrimination consistent with good testing practices and by the minimum number of parts necessary to produce sophisticated coverage of the testing problem.

In the summer and early fall of 1961 over 43,000 individual tests, including over 7,000 speaking tests, were administered, scored, and reported. The testing population included the 68 Summer and 7 Academic-Year NDEA Institutes. The test results were analyzed and the findings published by Educational Testing Service.

II. The Tests: The contractual obligation to develop two thirty-one test batteries in five languages (French, German, Italian, Russian, Spanish) covering the seven competencies (Listening Comprehension, Speaking, Reading, Writing, Applied Linguistics, Culture-Civilization, Professional Preparation) has been completed according to the time schedule delineated in the contract. What is more important is the *quality* of the tests and their value to the foreign-language profession and the educational community. At this point in time it is perhaps best to limit ourselves to what is statistically verifiable in answering the questions implicit in the preceding sentence. At the original conferences of the MLA chairmen and the ETS advisors to set the specifications for the tests it was agreed to aim at individual test reliability of .80 and battery reliability of .90 in accordance with accepted testing procedures. It may be clearly stated that in every case the minimum requirements for validity and reliability have been appreciably exceeded. Responsible officers of Educational Testing Service have stated that these MLA test batteries are among the most valid and reliable test batteries with which the ETS has ever been associated.

Where human scorers are concerned, as in the Speaking and Writing tests, one of the critical

problems is scorer reliability, since no tests can ever be more reliable in effective use than the scoring processes applied to them. In this context, we can report that, on the basis of scorer reliability checks for the Writing tests, we have achieved a figure of .996, which is about as good as can be obtained with machine scoring. A special problem has arisen in this regard in connection with the Speaking tests. In the first place, ETS has not computed the statistical reliability of the individual Speaking tests because every item on a given tape has been rated by a single rater and there is certain to be a "halo" effect because the scorer is agreeing with himself. All indications, however, are that we are obtaining high scorer reliability and plans are now in process to test the theory by having three different people score three different parts without hearing the other parts and with pre- and post-tests mixed in without the scorers' knowledge. A sample study (see Appendix C) for French shows a high reliability of .93 between the A and B forms and a correlation of .82 between pre- and post-tests. In addition, statistical analysis based on 1,336 cases shows a multiple correlation figure of .83 for the prediction of performance on the Speaking test from the scores of the Listening, Reading, and Writing tests. All indications, therefore, are that we are dealing with a reliable test in this new area of oral production measurement. A second problem that occurred particularly with the 1960 preliminary forms was the fact that a significant number of participants received lower scores on post-tests than on pre-tests. It is not possible to suppose retrogression on the part of the number of cases involved and our conclusion, after multiple checking of all statistics, was that the scorers became gradually more rigorous as scoring proceeded. This effect was practically eliminated in 1961 and we assume that we have the matter under control. The following table gives the reliability figures available at present for the individual tests.

	<i>Form A</i>	<i>Form B</i>
<i>French</i>		
Listening	.917	.912
Reading	.930	.930
Writing	.935	.942
Linguistics	.872	.863
Culture-Civilization	.858	.868
<i>Spanish</i>		
Listening	.911	.908
Reading	.913	.899
Writing	.944	.948
Linguistics	.837	.849
Culture-Civilization	.889	.886

MLA FL Proficiency Tests

<i>Italian</i>	<i>Form A</i>	<i>Form B</i>	<i>German (based on 297 cases)</i>		
Listening	.845	.862	Listening	R = .83	SEest = 4.90
Reading	.935	.867	Speaking	R = .83	SEest = 10.89
Writing	.965	.950	Reading	R = .90	SEest = 4.85
Linguistics	.845	.781	Writing	R = .90	SEest = 7.08
Culture-Civilization	.889	.829	<i>Russian (based on 176 cases)</i>		
<i>German</i>			Listening	R = .87	SEest = 3.26
Listening	.921	.890	Speaking	R = .84	SEest = 10.38
Reading	.942	.928	Reading	R = .80	SEest = 6.50
Writing	.968	.964	Writing	R = .81	SEest = 9.37
Linguistics	.898	.887			
Culture-Civilization	.882	.866			
<i>Russian</i>					
Listening	.815	.910			
Reading	.928	.919			
Writing	.962	.962			
Linguistics	.818	.855			
Culture-Civilization	.857	.845			
<i>Professional Preparation</i>	.861	.874			

Appendix C is a table which shows further reliability figures between forms and correlations for pre- and post-testing. The tests were carefully spiraled at all administrations to insure statistical validity.

Appendix D is a table which shows the inter-correlations between the competencies measured. It is interesting to note that there is a fairly high correlation between the four skills and indeed formulas have been derived which will permit with fair reliability prediction of any one skill performance from the other three. The correlations are not high enough, however, for us to recommend omission of any one test. In all cases the standard error is greatest for the Speaking and Writing tests and least for Listening and Reading. In the following table showing the predictability of performance by skills based on performance in the other three skills R is the correlation between predicted and observed values. SEest is the standard deviation of the difference between the predicted and observed values. The number of cases is too small in Italian to produce meaningful values.

French (based on 1,336 cases)

Listening	R = .85	SEest = 4.73
Speaking	R = .83	SEest = 11.72
Reading	R = .89	SEest = 5.02
Writing	R = .89	SEest = 6.03

Spanish (based on 1,334 cases)

Listening	R = .85	SEest = 4.10
Speaking	R = .78	SEest = 12.98
Reading	R = .89	SEest = 4.61
Writing	R = .90	SEest = 5.90

Our conclusion, therefore, is that the correlations are not high enough and the standard errors are too large for us to recommend the exclusion of any one test particularly when the reliability of the individual skill tests is so high as to insure reasonably accurate measurement of the individual skills in themselves of such interest to our profession at this time.

One trend indicated in Appendix D of particular interest is the relatively low correlation of the Professional Preparation test with the other competencies. Since the forms were spiraled and both were about equally involved in pre- and post-testing, the conclusion must be that this trend is not a function of the test. From the point of view of testing objectives this means that in this area we are measuring something quite different from the other areas, yet admittedly it is one of considerable significance to our profession.

The high level of reliability and statistical validity of the tests, a battery reliability of over .90 for each of the languages, is, in conclusion of this section, the point we take pleasure in underlining again. If we assume that the language profession is to concern itself with proficiency measurement, and all indications are that such is the case, then it may be said that project SAE 8349 has provided the sensitive instruments with which to make such measurement.

III. What the Tests Are Telling Us: The wide employment of the tests in the Summer and Academic-Year NDEA Institutes, in preliminary form to provide us with data for revision and in final form to provide us with normative data, has played a double role, for it furnishes the government and the profession with objective data which supports and justifies the Institute program as well as criteria for analyzing the effectiveness of individual Institutes. It is satisfying to note that the Institute population, which together with control groups, Indiana and Michigan Study Groups and the Associated Colleges of the Mid-West project provided 2,866 examinees and data from 40,124 tests, in the summer of 1961 made measurable progress

in all five languages and in all seven competencies without exception as illustrated in the following table.

Pre-Test and Post-Test Means Compared

French (converted scores)

	<i>Pre-</i>	<i>Post-</i>	<i>Gain</i>
Listening	M 39.0	42.3	3.3
Speaking	M 78.5	84.1	5.6
Reading	M 44.5	46.6	2.1
Writing	M 45.2	46.4	1.2
Applied Linguistics	M 45.5	50.7	5.2
Culture-Civilization	M 44.3	49.7	5.4
Professional Prep.	M 59.1	66.9	7.8

German (converted scores)

	<i>Pre-</i>	<i>Post-</i>	<i>Gain</i>
Listening	M 38.9	42.2	3.3
Speaking	M 81.2	84.5	3.3
Reading	M 45.3	48.7	3.4
Writing	M 46.3	49.0	2.7
Applied Linguistics	M 48.1	53.9	5.8
Culture-Civilization	M 48.6	53.9	5.3
Professional Prep.	M 60.2	67.3	7.1

Spanish (converted scores)

	<i>Pre-</i>	<i>Post-</i>	<i>Gain</i>
Listening	M 38.4	41.6	3.2
Speaking	M 73.1	77.5	4.4
Reading	M 41.5	44.1	2.6
Writing	M 45.2	48.7	3.5
Applied Linguistics	M 43.5	49.0	5.5
Culture-Civilization	M 49.5	56.3	6.8
Professional Prep.	M 58.8	66.8	8.0

Russian (converted scores)

	<i>Pre-</i>	<i>Post-</i>	<i>Gain</i>
Listening	M 39.2	43.1	3.9
Speaking	M 72.0	79.5	7.5
Reading	M 34.7	38.9	4.2
Writing	M 50.6	57.0	6.4
Applied Linguistics	M 44.2	48.6	4.4
Culture-Civilization	M 49.4	53.1	3.7
Professional Prep.	M 59.7	64.4	4.7

Italian (raw scores)

	<i>Pre-</i>	<i>Post-</i>	<i>Gain</i>
Listening	M 20.3	22.6	2.3
Speaking	M 63.3	71.0	7.4
Reading	M 26.8	30.6	3.8
Writing	M 31.0	34.4	3.4
Applied Linguistics	M 27.7	31.0	3.3
Culture-Civilization	M 30.0	33.5	3.5
Professional Prep.	M 33.0	41.3	8.3

Furthermore we have provided all data to measure individual participant scores against Institute and national scores and to measure the

individual Institute mean score gains against the national mean score gains. For norming purposes all data have been converted to percentile equivalents.

The above table illustrates some interesting trends. It indicates, for example, that the greatest mean gains throughout the Institutes were made in the areas of Applied Linguistics, Culture-Civilization, and Professional Preparation. It is unquestionably an indication of the fact that the population was least knowledgeable about these areas to begin with and hence could show more dramatic progress, but it is also significant in terms of the purposes of the Institutes, which are in part to emphasize these areas. In the non-skill areas, greatest gains were consistently made in the area of Professional Preparation. In this area, which we have observed not to have high correlation with the others, it can be stated that the Institute Program is having significant impact indeed. Of the skill tests, the Speaking competency generally shows the most appreciable gains, a fact which is also significant in terms of this major purpose of the Institute program. The fact that Writing generally shows least gain of the skills is probably an indication that it receives less emphasis than the other skills. A use of the tests is thus pointed up in that Directors will be able to decide on the basis of objective data where new emphases need to be placed.

In a recent administration to a native-speaking group, it was observed that although high scores were obtained in the skill areas, no advantage to the native speaker could be observed in the three non-skill areas. In fact the scores in these areas were significantly low. Assuming again the professional validity of the "qualifications" criteria which guided the construction of the tests, our conclusion may well be that native speakers need the Institute program or special programs emphasizing the non-skill areas as a part of their training for teaching foreign languages in this country.

IV. Impact and the Future: Easily two hundred individuals, as members of the test construction committees, as scorers,¹ as voices on tapes,

¹ The scorers, who were carefully selected through interviews for their linguistic training and skills, met together for orientation training at the beginning of each scoring session. The scorers for French were: Marcella Buxbaum, Carolyn Goldberg, Dennis Healy, José Huertas-Jourda, Lillian B. Jeanpierre, Wendell A. Jeanpierre, Fred Myers, Cecile Nebel Marie Louise Pesselier, Razel Pincus, Annette Schwartzberg, Carolyn Strauss, Marcel Wallace; for German: Carl Buchman, R. Travis Hardaway, Margaret Mong, Senta Stiefel; for Italian: Donatella Careccia, Marcia Cobourn; for Russian: Alexander Chlopoff, Irene Gendzier, George Holen-

and as consultants, have participated in the project. They represent all teaching ranks in many different institutions at all levels throughout the country. This fact alone has not only disseminated the purposes of the tests and information about them, but it has involved many people in proficiency testing as a function of the teaching process. The devotion and high quality of work which has characterized the efforts of all from the beginning is a testimonial to the sense of professionalism existing among language teachers.

In connection with his MLA duties, the Director has been called upon to write several articles, to give numerous speeches to professional groups, and to attend a number of conferences concerning testing in other disciplines. A list follows as an indication of the type of dissemination involved.

Articles

Illinois Educational Press Bulletin, December 1959

ETS Developments, VIII 1, October 1959

PMLA, May 1961

MLJ, in preparation

Speeches

Foreign Language Association of Northern California, Stanford University

ETS Invitational Conference, New York

Education Seminar, University of Maine

Annual Meeting MLA, Philadelphia

Central States Modern Language Teachers Association, Cleveland

Conferences addressed on the subject of the tests

The Council on Cooperation in Teacher Education, Washington, D. C.

The National Commission on Accrediting, Washington, D. C.

The National Council for Accreditation of Teacher Education, Washington, D. C.

Conference 22 (Teacher Training Curricula in the Foreign Language Field), MLA Annual Meetings, Chicago, Ill.

Institute Directors Meetings, Chicago and Boulder

Regional TEPS Conference, Boston

Cooperative Classroom Testing Project, Princeton

Center for Applied Linguistics Conference on Testing Common and Less Common Languages, Washington, D. C.

ETS-AGS Conference on Testing for Graduate Language Requirements, Princeton

Conferences on Testing English as a Foreign Language, Washington, D. C.

Post Hoc Ergo Propter Hoc? Whether or not there is a direct relationship between the foregoing activities and subsequent ones, it has been clear to the Director that there has been steadily increasing interest in the concept of proficiency testing. A case in point is the following quotation

from the annual report of the Council on Cooperation in Teacher Education in 1960. "The Council on Cooperation in Teacher Education recommends that research and development in proficiency examinations of all kinds be encouraged and subsidized. The Modern Language Association in proposing requirements and in developing tests for assessing competency in the teaching of foreign languages has set an example that may be followed by other academic and professional disciplines." The Director is also currently Chairman of the French Committee for the Association of Graduate Schools' project to develop standardized tests for use in graduate schools in connection with language requirements. He is serving on the National Advisory Council for a project on the Testing of English as a Foreign Language and is exploring, at the request of several government agencies, the problem of tests in common and less common languages. It seems evident to the Director that the MLA project, as a first of its kind in the foreign-language field, has created new interest in the development of standardized tests for the measurement of foreign-language skills through having demonstrated the fact that reliable instruments can be built which are consonant with the uses of foreign languages in today's world.

By far the greatest impact, however, derives from the numbers of people to whom the tests have been administered. By the end of the 1962-63 academic year we estimate that nearly 10,000 individuals will have been tested and more than 132,000 tests administered. The following table illustrates the uses of the tests to that date.

Figures on Institutes, Participants, and Tests

Summer 1960

37 Summer Institutes

2,154 Examinees (including Carnegie Inter-University Program and Middlebury Italian School)

28,625 Tests Administered

Academic Year 1960-61

5 Academic Year Institutes

108 Examinees

1,512 Tests Administered

Summer 1961

67 Summer Institutes

2,866 Examinees (including Indiana and Michigan Study Groups, and Associated Colleges of the Midwest)

koff, Rose Lefel, Natalia Sukacev; for Spanish: Ethel Arcilagos, Lucia Bonilla, Mary Cannizzo, Vincent Durkin, Victor Fuentes, Antonio Gila, Juan Lopez, Margaret McEvoy, Rizel Pincus, Mrs. Stanley Redka.

40,124 Tests Administered
 Academic Year 1961-62
 7 Academic Year Institutes
 162 Participants
 2,268 Tests Administered and Scored
 Summer 1962
 79 Summer Institutes
 4,028 Examinees
 56,392 Tests to be Administered
 Academic Year 1962-63
 5 Academic Year Institutes
 119 Examinees
 1,666 Tests to be Administered
 Miscellaneous Small Programs in which we are already
 involved
 11 Institutions or Agencies
 421 Number of Examinees to Date
 2,105 Number of Tests to Date

Grand Totals

	<i>Examinees</i>	<i>Tests Administered</i>
1960-62	5,290	72,529
1962-63	4,568	60,163
Total	9,858	132,692

Such widespread use cannot help but have an impact upon language teachers in terms of an increased sense of professionalism and in the profit which derives from the identification of strengths and weaknesses.

In addition, we have had over 200 requests from State and Local Boards of Education, from various institutions and agencies, from departments of foreign languages and from individuals. In anticipation of the ongoing interest, the Office of Education has this spring granted permission to the MLA for a ten-year period "to reproduce, administer, distribute in a manner consistent with test security, and otherwise to exploit in the public interest, the Foreign Language Proficiency Tests for Teachers and Advanced Students including any revisions thereof." The MLA has in consequence entered into a contract with ETS for a two-year period to insure their professional assistance and continuing collaboration in dis-

tribution, administration, and scoring of the tests as well as to secure their help in making revisions and new forms.

The Director is pleased to report that the following Institutions and Agencies are already using the tests or have contracted for their use: Hampton Institute (for undergraduate majors); University of Massachusetts; Emmanuel Missionary College, Michigan; Indiana University (Russian Study Group); Oberlin (French Study Group in France), (Spanish Study Group in Mexico); Associated Colleges of the Mid-West (Saint Olaf, Beloit, Monmouth, Lawrence, Coe); State of Pennsylvania; State of New Hampshire; State of Delaware; Washington D. C. (Local Board). In addition, the following States are seriously exploring the possibility: Vermont, Hawaii, Massachusetts.

Reports from the users indicate enthusiastic satisfaction and it seems reasonable to be optimistic about the gradual increase in participation. An eight-page brochure describing the tests in some detail is planned for fall distribution.

V. Concluding Remarks: No one is more aware than the Project Director that such success as may be ascribed to the completion of the project is due to the cooperation and competence of all those who have participated. He would be remiss, however, if he did not name for special mention George Winchester Stone, Jr., Donald D. Walsh, Allan Hubbell, and Harry Alonso of the MLA Staff, and Robert Solomon and Mrs. Miriam Bryan of ETS. The project could not have been completed without their constant support. The Director is embarrassed only by the fact that space does not allow for the listing of the many others to whom major credit is due in both organizations. He feels, too, that it must be stated again that the MLA-ETS collaboration has been most successful and mutually reinforcing. It has been an experience that is a tribute to the professional quality and attitudes of both organizations.

APPENDIX A. QUALIFICATIONS FOR TEACHERS OF MODERN FOREIGN LANGUAGES

COMPETENCE	SUPERIOR	GOOD	MINIMAL
Listening Comprehension	Ability to follow closely and with ease all types of standard speech, such as rapid or group conversation and mechanically transmitted speech.	Ability to understand conversation of normal tempo, lectures, and news broadcasts.	Ability to get the sense of what an educated native says when he is making a special effort to be understood and when he is speaking on a general and familiar subject.
Speaking	Ability to speak fluently, approximating native speech in vocabulary, intonation, and pronunciation. Ability to exchange ideas and to be at ease in social situations.	Ability to talk with a native without making glaring mistakes, and with a command of vocabulary and syntax sufficient to express one's thoughts in conversation at normal speed with reasonably good pronunciation.	Ability to read aloud and to talk on prepared topics (e.g., for classroom situations) without obvious faltering, and to use the common expressions needed for getting around in the foreign country, speaking with a pronunciation understandable to a native.
Reading	Ability to read almost as easily as in English material of considerable difficulty.	Ability to read with immediate comprehension prose and verse of average difficulty and mature content.	Ability to grasp directly (i.e. without translating) the meaning of simple, non-technical prose, except for an occasional word.
Writing	Ability to write on a variety of subjects with idiomatic naturalness, ease of expression, and some feeling for the style of the language.	Ability to write a simple "free composition" such as a letter, with clarity and correctness in vocabulary, idiom, and syntax.	Ability to write correctly sentences or paragraphs such as would be developed orally for classroom situations and to write a simple description or message without glaring errors.
Applied Linguistics	The "good" level of competency with additional knowledge of descriptive, comparative, and historical linguistics.	The "minimal" level of competency with additional knowledge of the development and present characteristics of the language.	Ability to apply to language teaching an understanding of the differences in the sound system, forms, and structures of the foreign language and English.
Culture and Civilization	An enlightened understanding of the foreign people and their culture, such as is achieved through personal contact, through travel and residence abroad, through study of systematic descriptions of the foreign culture, and through study of literature and the arts.	The "minimal" level of competency with first-hand knowledge of some literary masterpieces and acquaintance with the geography, history, art, social customs, and contemporary civilization of the foreign people.	An awareness of language as an essential element of culture and an understanding of the principal ways in which the foreign culture differs from our own.
Professional Preparation	A mastery of recognized teaching methods, evidence of breadth and depth of professional outlook, and the ability to experiment with and evaluate new methods and techniques.	"Minimal" level of competency plus knowledge of the use of specialized techniques, such as audiovisual aids, and of the relation of language teaching to other areas of the curriculum. Ability to evaluate the professional literature of foreign language teaching.	Knowledge of the present-day objectives of the teaching of foreign languages as communication and an understanding of the methods and techniques for attaining these objectives.

N.B. The names of the seven competencies were also slightly modified and appear in the test batteries as listed here.

APPENDIX B. TEST CONSTRUCTION COMMITTEES

LISTENING COMPREHENSION

French	Edward Geary, Chairman	Harvard
	Alain Seznec	Cornell
	Edmond Méras	Phillips Exeter
Spanish	Patricia O'Connor, Chairman	Brown
	Sol Saporta	Washington
	Filomena Peloro	Mat. Dev. Center, NYC
Italian	James Ferrigno, Chairman	Massachusetts
	Carlo Vacca	Wellesley (Mass.) H.S.
	Rigo Mignani	Harpur
German	Jack Stein, Chairman	Harvard
	Walter Lohnes	Phillips Academy
	Hugo Schmidt	Bryn Mawr
Russian	Richard Burgi, Chairman (1959-60)	Yale
	Rostislav Rozdestvensky (1959-60),	Glastonbury (Conn.) Sch.
	Chairman (1960-61)	
	Nina Berberova-Kochevitsky	New Haven, Conn.

SPEAKING

French	James Iannucci, Chairman	St. Josephs
	Frederic St. Aubyn	Delaware
	Annette Emgarth	Dover, Delaware
Spanish	Stanley Sapon, Chairman	Ohio State
	Edward Allen	Univ. School, Ohio State
	Chris Nacci (1959-60)	Capital
Italian	Sandra Scharff (1960-61)	Ohio State
	Robert Politzer, Chairman	Michigan
	Peter Fodale	Michigan
German	Fred Bosco	Michigan
	Herbert Penzl, Chairman	Michigan
	Mary Crichton	Michigan
Russian	Max Dufner	Michigan
	William Edgerton, Chairman (1959-60)	Indiana
	Horace Dewey (1959-61), Chairman (1960-61)	Michigan
	Nonna Shaw	Indiana

READING

French	Linn Edsall, Chairman	Wayne State
	Jane Bourque	Madison, Conn.
	Philip Wadsworth (1959-60)	Illinois
Spanish	Paula Thibault (1960-61)	Detroit, Mich.
	Frederick Agard, Chairman (1959-60)	Cornell
	Dalai Brenes (1959-61), Chairman (1960-61)	Cornell
Italian	Katherine Whitmore	Smith
	Norma Fornaciari, Chairman (1959-60)	Roosevelt
	Clarence Turner (1959-61), Chairman	
German	(1960-61)	Rutgers
	Maria Piccirilli	Vassar
	Guido Guarino (1960-61)	Rutgers
	C. R. Goedsche, Chairman	Northwestern
Russian	Werner Hollman	Princeton
	Meno Spann	Northwestern
	Assya Humesky, Chairman	Syracuse
	Horace Dewey (1959-60)	Michigan
	Dale Winkels (1959-60)	Michigan
	Clayton Dawson (1960-61)	Syracuse
	Nicholas Karateew (1960-61)	Syracuse

MLA FL Proficiency Tests

WRITING

French	Nelson Brooks, Chairman	Yale
	Pierre Capretz	Yale
	Gordon Christopher	Hillhouse H.S., New Haven
Spanish	Elizabeth Nicholas, Chairman	Mat. Dev. Center, NYC
	Jeannette Atkins	Staples H.S., Westport, Conn.
	Jaime Muirden	New Haven, Conn.
Italian	Robert Serafino, Chairman	State Dept. of Ed., Conn.
	Bianca Calabresi	Albertus Magnus
	Arthur Selvi	Central Conn. State Coll.
German	Joseph Reichard, Chairman	Oberlin
	Edith Runge	Mount Holyoke
	Walter Lohnes	Phillips Andover
Russian	Horace Lunt, Chairman (1959-60)	Harvard
	Bayara Tschirwa (1959-61), Chairman	Harvard
	(1960-61)	
	Dmitry Grigorieff	Columbia
	Marina Prochoroff (1960-61)	Mat. Dev. Center, NYC

APPLIED LINGUISTICS

French	Robert Politzer, Chairman	Michigan
	Harry Bratnober	Macalester
	Albert Vladman	Indiana
	Fernand Marty	Hollins College
Spanish	Sol Saporta, Chairman	Washington
	Patricia O'Connor	Brown
	Mary Temperly (1960-61)	Illinois
	Ismael Silva-Fuenzalida (1959-60)	Foreign Service Institute School of Languages
Italian	Edward Williamson, Chairman	Wesleyan
	Anthony Pellegrini	Vassar
	Salvatore Castiglione (1959-60)	Georgetown
	Ernest Pulgram (1960-61)	Michigan
German	Freeman Twaddell, Chairman	Brown
	R. M. S. Heffner	Wisconsin
	William Moulton	Princeton
Russian	William Cornyn, Chairman	Yale
	Vladmir Petrov	Yale
	Howard Garey	Yale

CULTURE-CIVILIZATION

French	Georges May, Chairman	Yale
	Joseph Stookins	Loomis
	Kenneth Cornell	Yale
Spanish	Theodore Andersson, Chairman	Texas
	Miguel Enguñados	Texas
	Andrea McHenry	Houston Schools
Italian	Charles Speroni, Chairman	UCLA
	Aldo Scaglione	UC (Berkeley)
	Gaetano Pomposo	Pittsburgh (Calif.) H.S.
German	Else Fleissner, Chairman	Wells
	Karl Koenig	Colgate
	Anthony Schepsis	Utica Free Academy
Russian	Leon Stilman, Chairman (1959-60)	Columbia
	William Harkins, Chairman (1960-61)	Columbia
	Peter Juviler (1959-60)	Hunter
	Mrs. Edward C. Bill (1959-60)	Princeton
	Henry Morton (1960-61)	Queens
	Francis Randall (1960-61)	Columbia

PROFESSIONAL PREPARATION

Alfred Pellegrino, Chairman
 Germaine Cressey
 Mary Thompson
 George Scherer
 Emma Birkmaier
 Meyer Krakowski (1959-60)

(Italian)
 (French)
 (Spanish)
 (German)
 (Russian)

Maine
 Montclair
 Glastonbury, Conn.
 Colorado
 Minnesota
 L. A. City College

APPENDIX C. RELIABILITY OF MLA FOREIGN LANGUAGE PROFICIENCY TESTS

BETWEEN A AND B FORMS			PRE- AND POST- TEST CORRELATIONS		BETWEEN A AND B FORMS			PRE- AND POST- TEST CORRELATIONS	
FRENCH					Writing	.95		.90	
Listening	.91		.84	App. Ling.	.84		.78		
Speaking	.93		.82	Cult.-Civ.	.89		.78		
Reading	.94		.88	Prof. Prep.			.70		
Writing	.87		.78						
Cult.-Civ.	.86		.74					ITALIAN	
Prof. Prep.			.67	Listening	.85		*		
GERMAN					Speaking				
Listening	.90		.84	Reading	.90				
Speaking			.81	Writing	.96				
Reading	.94		.90	Cult.-Civ.	.81				
Writing	.97		.92	Prof. Prep.	.86				
App. Ling.	.89		.85					RUSSIAN	
Cult.-Civ.	.87		.78	Listening	.86		.80		
Prof. Prep.			.64	Speaking			.79		
SPANISH					Reading	.92		.85	
Listening	.90		.86	Writing	.96		.91		
Speaking			.79	App. Ling.	.84		.77		
Reading	.90		.88	Cult.-Civ.	.85		.75		
				Prof. Prep.			.67		

* Not estimated because of limited number of cases.

MLA FL Proficiency Tests

APPENDIX D. INTERCORRELATIONS BETWEEN SKILLS

FRENCH							
	LISTENING	SPEAKING	READING	WRITING	LING.	CULT.-CIV.	PROF. PREP.
Listening		.784	.800	.782	.566	.540	.317
Speaking	.784		.739	.804	.523	.487	.269
Reading	.800	.739		.859	.638	.643	.371
Writing	.781	.804	.858		.660	.564	.344
Linguistics	.566	.523	.638	.660		.592	.548
Cult.-Civ.	.540	.487	.634	.564	.592		.477
Prof. Prep.	.317	.279	.371	.344	.548	.477	
GERMAN							
Listening		.727	.817	.781	.509	.512	.230
Speaking	.727		.742	.787	.551	.403	.206
Reading	.817	.742		.860	.624	.614	.274
Writing	.781	.787	.860		.738	.629	.326
Linguistics	.509	.551	.624	.738		.670	.460
Cult.-Civ.	.512	.403	.614	.629	.670		.455
Prof. Prep.	.230	.206	.274	.326	.460	.455	
SPANISH							
Listening		.925	.797	.796	.479	.599	.371
Speaking	.925		.691	.724	.412	.525	.293
Reading	.797	.691		.857	.568	.687	.430
Writing	.796	.724	.857		.625	.677	.467
Linguistics	.479	.412	.568	.625		.589	.616
Cult.-Civ.	.599	.525	.687	.677	.589		.520
Prof. Prep.	.371	.293	.430	.467	.616	.520	
ITALIAN							
(Intercorrelations Impossible Because of Small Number of Cases)							
RUSSIAN							
Listening		.743	.770	.746	.480	.408	.238
Speaking	.743		.697	.727	.436	.304	.160
Reading	.770	.697		.669	.367	.328	.078
Writing	.746	.727	.669		.708	.480	.298
Linguistics	.480	.436	.367	.708		.506	.511
Cult.-Civ.	.408	.304	.328	.480	.506		.408
Prof. Prep.	.238	.160	.078	.298	.511	.408	